How socio-indexical information modulates the relationship between formant variability and vowel categorization

Nhung Nguyen[1], Jason A. Shaw[1,2], Catherine T. Best[1,2], and Michael D. Tyler[1,3]
[1] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
[2] School of Humanities and Communication Arts, Western Sydney University, Australia
[3] School of Social Sciences and Psychology, Western Sydney University, Australia
nhung.nguyen@westernsydney.edu.au

Phonological representations have recently been viewed as probability distributions over phonetic parameters (Kleinschmidt & Jaeger, 2015; Norris & McQueen, 2008; Pierrehumbert, 2001, 2003, 2006). In this view phonetic variability is central to phonological representations and crucial to predicting speech perception behaviour. A general prediction is that phonological categories with wider distributions over a phonetic parameter are more tolerant of variation in that phonetic dimension than categories with narrower distributions. Estimates of this variability in representative speech corpora may approximate phonological representations, but there is evidence that listeners shift their expectations about the relationship between phonetic parameters and phonological categories dynamically across socio-indexical dimensions (Hay & Drager, 2010; Hay et al., 2006). Following Kleinschmidt and Jaeger (2015), we hypothesize that listeners have multiple distributions for a vowel that correspond to different socio-indexical dimensions (e.g., regional accent, age, and social class). In Australian English, the vowels FLEECE, NEAR, KIT, DRESS, SQUARE, GOOSE, and NURSE typically carry such social information (Cox, 1999, 2006a; Harrington et al., 1997). For these vowels, corpus-based estimates of phonetic parameter variability may be inappropriate estimates of the distributions used by listeners in perception, as listeners may quickly adapt to the speaker and apply a socially appropriate (and accordingly narrow) phonetic distribution. Thus, the general prediction that wide distributions (based on corpus measurements) lead to increased tolerance of variability in perception may only apply to relatively stable vowel categories (i.e., vowel categories that do not vary systematically according to socio-indexical dimensions).

In this study, we tested the hypotheses that vowel representations are distributions over formant values and that formant variability is interpreted differently for vowels carrying socio-indexical information than for more stable vowels. We estimated distributions over formant values of 13 Australian English monophthongs from a /hVd/ corpus of 60 Sydney Northern Beaches speakers (Cox, 2006), and then explored the relationship between formant variability and the performance of 64 listeners in a vowel categorization task in which listeners labelled the vowels in /hVdə/ stimuli using native vowel categories. We fitted binomial mixed effects models to the accuracy data (3,328 data points) in R and carried out model comparisons. Participant and vowel token were included as random effects. Fixed effects were the socio-indexicality of the vowels, which was a binary variable dividing the typically socio-indexical vowels (i.e., FLEECE, NEAR, KIT, DRESS, SQUARE, GOOSE, and NURSE) from the typically non-socio-indexical ones (i.e., TRAP, STRUT, START, LOT, NORTH, and FOOT), and formant variability. Below are the main models:

- Data = 13 monophthongs:
  (1a) Accuracy ~ formant variability + vowel socio-indexicality + (1|Participant) + (1|Token)
  (1b) Accuracy ~ formant variability * vowel socio-indexicality + (1|Participant) + (1|Token)
- Data = typically non-socio-indexical subset:
  (2) Accuracy ~ formant variability + (1|Participant) + (1|Token)
- Data = typically socio-indexical subset:
  (3) Accuracy ~ formant variability + (1|Participant) + (1|Token)

Figure 1 shows an interaction between formant variability and the socio-indexicality of the vowels. This interaction was significant ($\chi = 8.49$, $p = .004$) by means of model comparison (Table 1). Results from models 2 and 3 confirm that the correlations between formant variability and categorization accuracy for both the socio-indexicality subsets are also significant (typically non-socio-indexical: $\beta = 0.01$, $p = .046$; typically socio-indexical: $\beta = -0.03$, $p = .02$). There is a positive correlation between formant variability and categorization accuracy, but only when the vowel means are stable across speakers. The positive relationship between formant variability and categorization accuracy for

vowels TRAP, STRUT, START, LOT, NORTH, and FOOT supports the hypothesis that phonological representations involve distributions over phonetic values. The opposite relationship between formant variability and categorization accuracy for vowels FLEECE, NEAR, KIT, DRESS, SQUARE, GOOSE, and NURSE may indicate that listeners used more specific and socially relevant distributions for these vowels in the perception task. This suggests that formant variability for typically socio-indexical vowels could be interpreted as separate phonetic distributions corresponding to specific socio-indexical dimensions rather than a single wide or narrow distribution as for typically non-socio-indexical vowels.

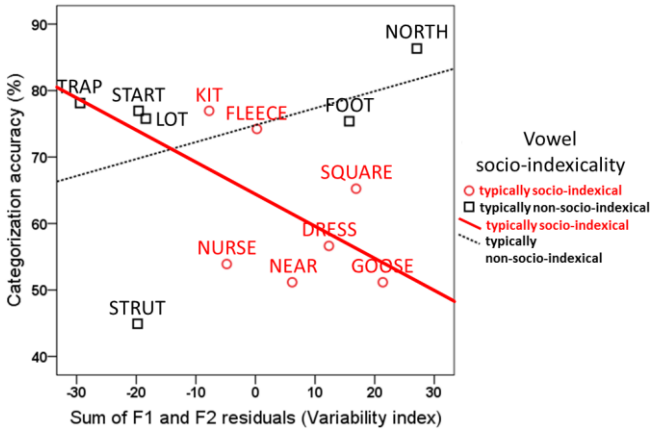

Figure 1: The different patterns of relationship between formant variability and categorization accuracy, depending on the socio-indexicality of the vowels. Red indicates the vowels that typically carry socio-indexical information. On the variability index, vowels with large positive numbers mean they are highly variable, and those with large negative numbers mean they are highly stable (Nguyen & Shaw, 2014).

Table 1: Variances explained by interaction model versus non-interaction model.

Model	AIC	BIC	logLik	deviance	Pr(>Chisq)
(1a) without interaction: Accuracy ~ formant variability + vowel socio-indexicality + (1|Participant) + (1|Token)	3686.0	3716.5	-1838.0	3676.0	
(1b) with interaction: Accuracy ~ formant variability * vowel socio-indexicality + (1|Participant) + (1|Token)	3679.5	3716.1	-1833.7	3667.5	.004

REFERENCES

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*, 1–27.

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179.

Cox, F. (2006a). Australian English pronunciation into the 21st century. *Prospect*, *21*(1), 3–21.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English. *Australian Journal of Linguistics*, *17*(2), 155–84.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865–892.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Nguyen, N., & Shaw, J. A. (2014). Why the SQUARE vowel is the most variable in Sydney. In *Proceedings of the 15th Australasian International Speech Science and Technology Conference*, 37–40.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*(6), 516–530.